Neural Style Transfer for Understanding How Stylistic Choices Affect Emotional Affect

Alex Weitzman
MS in Computer Science
Stanford University
alexwe@stanford.edu

Phoebe Yao
BS in Human-Centered Design Engineering
Stanford University
phoebeyao@gmail.com

Abstract

We used Multi-Style Neural Style Transfer (NST) on landscape paintings to explore how painting style effects the emotions of viewers. We grouped landscape paintings into "positive", "negative", and "neutral" valence categories and used groups of "positive" and "negative" valence style images in a Multi-Style NST to change the emotional affect of "neutral" valence content landscape painting. We also explored optimization methods such as color preservation, average pooling, and content-style trade-off to improve the appearance of our results. To investigate the performance of our algorithm, we ran a qualitative, randomized survey with 22 participants. As a result, we concluded that even while the emotional influence of art is subjective and controversial, NST can be effectively used to change the emotional affect of a painting, as well as to help identify common patterns in style that might lead to certain emotional responses from paintings.

1 Introduction

Paintings are the apex of human emotional expression. From portraying abstract dreams-capes to modern city life, visual artists aim to evoke particular feelings from their viewers. In addition to the content itself, one tool used to evoke specific emotions from a painting is style. Manipulations of elements such as color and brush stroke thickness can have a dramatic effect on the emotional response incited by a painting. In this way, even a scene that maintains its content may evoke a different emotion when depicted in a slightly different style. We aim to manipulate the style of landscape paintings to convey specific alternate emotional characteristics. We will accomplish this by using Neural Style Transfer (NST) and Transfer Learning to build upon a previously trained convolutional network. We will use the VGG-19 network trained on the ImageNet database. Along with this, we have access to a dataset of 2813 paintings scraped from the Wikiart dataset that have been tagged with discrete emotions by human labelers. The content image that would be inputted to our NST algorithm would be a painting that has been labeled "neutral" in our dataset. We would then take a group of style reference images, selected from our dataset as paintings with similar styles and artists, who represent the same emotional affect (for example, a group of paintings). will run NST with multiple style reference images that encompass a specific emotion. In doing so, our goal is to later analyze and identify patterns in the generated paintings to see what common elements of style lead to given emotions. This is interesting because we don't know the exact factors that contribute to a perceived emotion from art, so the results of such a style transfer are unpredictable. Since emotions are highly subjective, the results might be controversial and thought-provoking, raising more abstract questions about art and its meaning.

CS230: Deep Learning, Winter 2018, Stanford University, CA. (LateX template borrowed from NIPS 2017.)

2 Related work

The most well-known fundamental paper on Neural Style Transfer was written by Gatys et al. in 2015. The paper outlined a novel way to change the style of images based on a content image and a style image. In the past, changing the style of an image required image analogy - one would need to train a pair of images, including a photo and an artwork depicting that photo, before being able to create new artwork from that photo, as it required learning the transformation from the original photo to the stylized photo. Instead, NST allowed to use only one content image and an unrelated style image to change the style of the content image. The Gatys paper leveraged a convolutional neural network (CNN) VGG-19 architecture that had been pre-trained to perform object recognition using the dataset from ImageNet. More recent work on Neural Style Transfer has explored variations and optimizations of the traditional model. For example, our project has found inspiration from papers which use multiple style reference images on a single content image rather than only one style, to generate a unique combination of styles (Wynen et al., 2018; Huang Belongie, 2017; Dumoulin et al., 2017.) Another work which contributed to our project was also by Gathys et al., and explored simple linear methods to preserve the color of a content image while transferring styles (2016). Finally, in the future we would like to try photorealism regularization from Luan et al. (2017): this paper discussed a method that prevents exaggerated image distortions from NST by constraining the generated image to be represented by locally affine color transformations of the input.

3 Dataset and Features

Our original dataset included 2813 images scraped from Wikiart.org that were tagged with 10 discrete emotions: joy, optimism, fear, surprise, lust, anger, sadness, neutral, envy, and love. We received the dataset from our mentor, Chez Mana, who is sponsoring the project idea with the approval of CS230 TAs. The people who tagged the images both have university degrees, one an artist and the other a computer programmer. They concentrated on the primary emotion of every artwork.

1	URL	Style	Artist	Genre	Emotion	Image URL	Title	Date
1334	https://www.wikia	Surrealism	M.C. Escher	cityscape	Neutral	https://uploads3.wikiart.org/images/m-c-escher/nonza-corsica.jpg	Nonza, Corsica	1934
1335	https://www.wikia	Impressionism	Konstantin Korovin	genre painting	Joy	https://uploads0.wikiart.org/images/konstantin-korovin/north-idilliya-1	North idilliya	1886
1336	https://www.wikia	Surrealism	M.C. Escher	cityscape	Optimism	https://uploads0.wikiart.org/images/m-c-escher/not_detected_20467	NOT DETECTED	1931
1337	https://www.wikia	Abstract Express	Mark Tobey	abstract	Neutral	https://uploads4.wikiart.org/images/mark-tobey/not-identified.jpg	not identified	
1338	https://www.wikia	Tonalism	James McNeill Whistler	cityscape	Neutral	https://uploads6.wikiart.org/images/james-mcneill-whistler/note-in-pi	Note in Pink and	1880
1339	https://www.wikia	Post-Impressioni	Maurice Prendergast	cityscape	Neutral	https://uploads1.wikiart.org/images/maurice-prendergast/notre-dame	Notre Dame	1907
1340	https://www.wikia	Impressionism	Francis Picabia	cityscape	Neutral	https://uploads2.wikiart.org/images/francis-picabia/notre-dame-the-e	Notre Dame, the	1906
1341	https://www.wikia	Art Nouveau (Mo	Koloman Moser	genre painting	Optimism	https://uploads7.wikiart.org/images/koloman-moser/november-1902.	November	1902
1342	https://www.wikia	Art Nouveau (Mo	Zinaida Serebriakova	portrait	Love	https://uploads3.wikiart.org/images/zinaida-serebriakova/nurse-with-	Nurse with baby	1912
1343	https://www.wikia	Expressionism	Paula Modersohn-Becker	sketch and study	Neutral	https://uploads5.wikiart.org/images/paula-modersohn-becker/nursing	Nursing Mother a	1902
1344	https://www.wikia	Realism	Ivan Shishkin	sketch and study	Neutral	https://uploads2.wikiart.org/images/ivan-shishkin/oak-on-the-shore-o	Oak on the shore	1857
1345	https://www.wikia	Realism	Fyodor Vasilyev	landscape	Neutral	https://uploads8.wikiart.org/images/fyodor-vasilyev/oaks.jpg	Oaks	
1346	https://www.wikia	Realism	Valentin Serov	sketch and study	Neutral	https://uploads7.wikiart.org/images/valentin-serov/october-1898.jpg!l	October	1898
1347	https://www.wikia	Realism	Valentin Serov	landscape	Neutral	https://uploads1.wikiart.org/images/valentin-serov/october-domotcan	October. Domotc	1895
1348	https://www.wikia	Early Renaissan	Paolo Uccello	religious painting	Love	https://uploads1.wikiart.org/images/paolo-uccello/oculus-depicting-th	Oculus depicting	1443
1349	https://www.wikia	Romanticism	Ivan Aivazovsky	cityscape	Neutral	https://uploads7.wikiart.org/images/ivan-aivazovsky/odessa-at-night-	Odessa at night	1846
1350	https://www.wikia	Realism	Mary Cassatt	genre painting	Neutral	https://uploads8.wikiart.org/images/mary-cassatt/offering-the-panel-t	Offering the Pane	1873
1351	https://www.wikia	Baroque	Gerrit Dou	genre painting	optimism	https://uploads5.wikiart.org/images/gerrit-dou/officer-of-the-marksma	Officer of the Ma	1630
1352	https://www.wikia	Realism	Vasily Polenov	landscape	Optimism	https://uploads1.wikiart.org/images/vasily-polenov/oka-near-tarusa.jp	Oka near Tarusa	
1353	https://www.wikia	Realism	Vasily Polenov	landscape	Joy	https://uploads2.wikiart.org/images/vasily-polenov/oka-river-vladimir-	Oka river. Vladim	1926
1354	https://www.wikia	Realism	Vasily Polenov	landscape	Neutral	https://uploads1.wikiart.org/images/vasily-polenov/oka-valley-1902.jp	Oka Valley	1902
1355	https://www.wikia	Realism	Vasily Polenov	landscape	Neutral	https://uploads5.wikiart.org/images/vasily-polenov/oka-summer.jpg!L	Oka. Summer.	

Figure 1: Wikiart Dataset, available at http://bit.ly/wikiart230

When choosing which paintings to use for our experiment, we wanted to be very intentional in how we controlled for different elements of the paintings. Our goal for this project was to determine how painting style affects the emotional impact of a painting. Therefore, we chose to focus on a genre of paintings in which content didn't have as much of an emotional influence on the observer as style. For example, portraits could have caused our dataset labelers to feel sad if the person in the painting was frowning. However, Neural Style Transfer would not change the content of our content image to cause someone to frown since it is determined by correlations between channels, so we would gain very little insight from using NST with sad portraits as a style reference (and nothing guarantees that these "sad" portraits had a "sad" style as well). Therefore, we focused on landscapes - this was both a great choice as we believe landscapes mainly affect the viewer based on stylistic choices such as color and painting stroke, and also because we had 1620 landscape paintings to choose from in our dataset. After narrowing it down to this category, we also tried to control for the artist and painting style (realism, impressionism, etc.) as much as possible, so that the differences in style across paintings would be attributed to differences in emotion rather than variables like era or artist.

Since emotions are highly subjective, we saw a lot of overlap and potential confusion in the labeling of different emotions: for example, love, optimism, and joy were all emotions that could be interpreted from the same painting, but our dataset only recorded one emotion per painting. Thus, we decided to generalize the emotional labels from our dataset to valence - positive, negative, or neutral - rather than specific emotions. Still, our general focus was on paintings formerly labeled "joy" vs. "sadness", as we felt those would be most interesting and objectively "positive" vs. "negative".

For the NST implementation, we used an open source Git repository implementing the NST methods outlined in Gatys et al (2015). For preprocessing, the implementation we used centered the pixel values by subtracting the mean value from the pixel values in the input image. Our algorithm also normalized the style layer weights. Our algorithm allowed for varying resolutions for the input images, but most of our input images were at least 512x512 pixels, as we read that style images don't provide enough style information below that size for effective results. Since Neural Style Transfer does not depend on a large dataset to be successful (as it runs on one content image and a few style images), we did not use data augmentation.

We used the feature space provided by the 16 convolutional and 5 pooling layers of the 19-layer VGG Network. We did not use any of the fully connected layers. The networks were normalized by scaling the weights so that the mean activation of each convolutional filter over images and positions was equal to one. Also, inspired by the Gathys paper, we started out with maximum pooling but found, similarly, that average pooling resulted in more appealing generated images.

4 Methods

Per standard NST models, our training focuses on minimizing the following loss function:

$$L = \left\| Content_C - Content_G \right\|_2^2 + \left\| Style_S - Style_G \right\|_2^2$$

This function aims to minimize the difference in content between the generated image and the content image, as well as the difference in style between the generated image and the reference style image. In our case, we augmented the above loss function to train our Multi-Style NST by taking the sum of the loss over n-chosen style images. In addition to our loss, we explored different hyper-parameters, including:

Average pooling: We modify the standard max-pooling layers in order to improve the gradient flow in the CNN. Average pooling results in more aesthetic and natural appearing images. Since max pooling takes the max value among pixel values, generated images typically result in more salient features.

Content-Style Trade-off: We vary the ratio between the emphasis on the content and style loss (/) as an extension to the loss function. We find a ratio of 5e-2 content to style worked the best on landscape paintings. This also appears to be the standard ratio in NST literature.

5 Experiments

Our methodology involved choosing a group of around ten landscape paintings from our labeled data set that shared the same valence, using that to represent the "style" of landscape paintings of that valence. After this, we ran Multi-Style NST twice on a neutral content image, once using a chosen group of 8 positive valence images and once using a chosen group of 8 negative valence images. We controlled for confounding variables by designing the twin style image groups to be analogous in artist, time-period, and genre of images. We believe the variable consistency allows for a more qualitatively balanced group comparison while also representing diverse colors, themes, and seasons in landscape paintings. Then we compared the results to see what style features stood out for each valence group, and we evaluated our model's performance by surveying observers in a randomized test to measure the emotional valence they felt from observing the generated images.

For our pilot test, we first trained our model on paintings by Albert Bierstadt, a luminist painter. We chose two of his landscape paintings labeled as "negative" valence for one style group, and two labeled "positive" as another style group, and transferred the style on one of his neutral paintings, once using the positive valence style group and once with the negative valence.

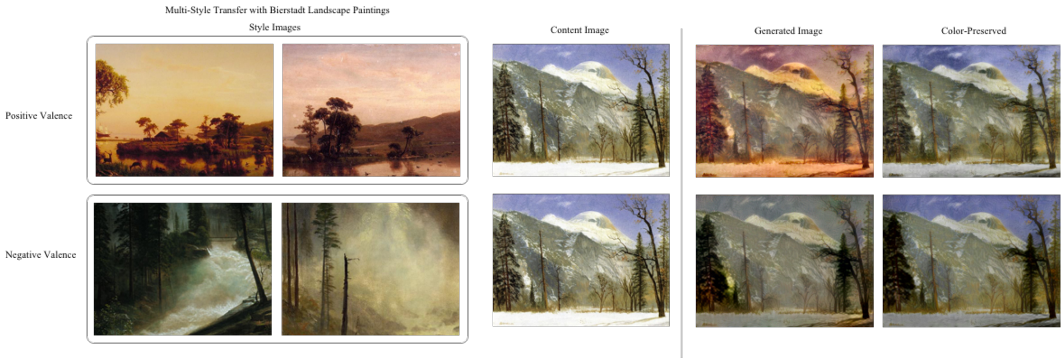

Figure 2:

We experimented and optimized for a variety of hyper-parameters. For one, we tested iteration sizes of 1000 and 5000 and found little difference in the total cost of the generated images. We also tested for color preservation, and total variation de-noising.

Positive Valence Generated Images

iteration: 1000
content loss: 290151
style loss: 1.44825e+06

iteration: 5000
content loss: 330623
style loss: 1.45498e+06

Figure 3: The top row displays a positive emotional transfer on a neutral content painting - before and after color preservation. The bottom displays a generated painting of negative valence.

To understand how the multi-style representation correlated between the different features in different layers of the CNN, we reconstructed the styles of the input images by training using solely the style loss. From left to right on figure 3, the subsets were: 1) ‘conv1 1’ and ‘conv2 1’, 2) ‘conv1 1’, ‘conv2 1’ and ‘conv3 1’, 3) ‘conv1 1’, ‘conv2 1’, ‘conv3 1’ and ‘conv4 1’, 4) ‘conv1 1’, ‘conv2 1’, ‘conv3 1’, ‘conv4 1’ and ‘conv5 1’. This creates images that match the style of the given set of images with finer features being represented in the earlier layers and coarser, larger scale features showing up in the later CNN layers. In comparing the positive and negative valence style reconstructions, we observe that the positive valence multi-style images tend towards lighter more colorful patterns, while the negative valence multi-style images are generally darker and higher contrast.

6 Conclusion/Future Work

Looking at our generated images for positive and negative valence, we noticed a few interesting trends. First, we saw that positive images tended to be more colorful. They seemed to have more of a glow, and when as we increased the number of iterations these elements stood out even more. In paintings of negative valence, light generally grew darker, and features of elements in the painting such as trees lost some of their detail and glow.

Still, after running our analysis, we found that much of the emotion captured by the painting was controversial. Not only was did viewers have controversial opinions of the emotion for our generated

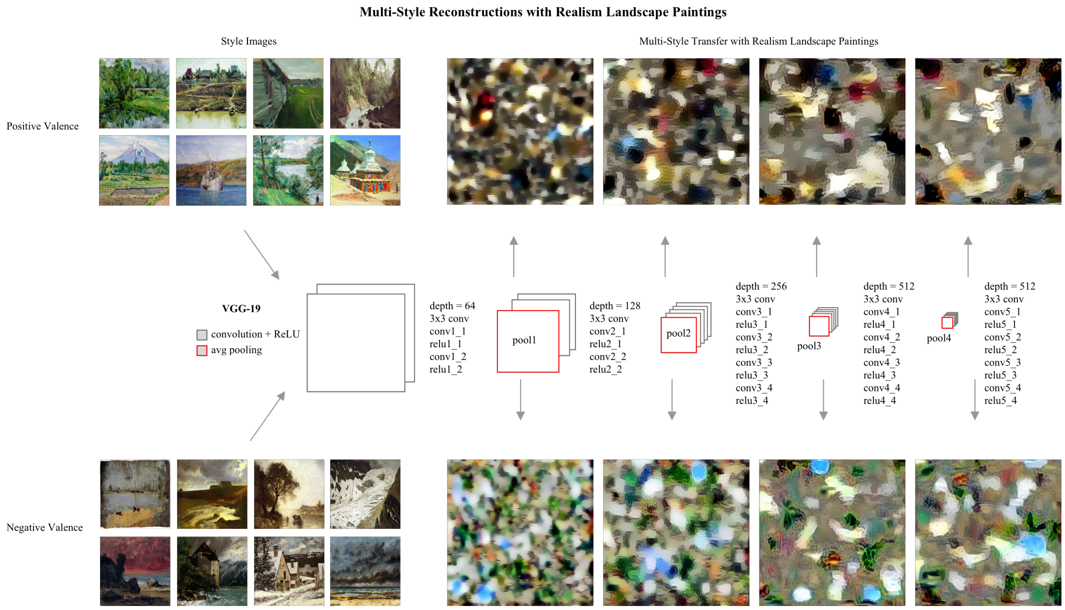

Figure 4: This figure shows a positive valence generated image comparing 1000 vs. 5000 iterations.

images, but also for the original images themselves that had already been labeled "neutral" in our database. It is impossible to predict what a human will feel when they witness a painting, and that's part of the magic and mystery of art. Furthermore, art is very complex, influenced by the background of the artist, the style, the symbolism, the viewer's knowledge, and the era when it was made. Finally, style is not all there is to the emotion captured by paintings - some paintings show content that might change a person's feelings, and Neural Style Transfer is not powerful enough to control that. In future work, we would love to explore ways in which the content of a painting might be altered to impact emotional effect. We would love to look into Generative Adversarial Networks and tackle the challenge of classifying the emotion of a painting based on content, using that as a discriminator while we generate different paintings to try to teach a model how to create paintings that capture a certain valence or emotion.

7 Contributions

Alex and Phoebe both worked on getting an initial Neural Style Transfer algorithm working. Phoebe then took the lead on transferring to a new implementation for the algorithm and on training the data, as well as optimization and exploring new experiments. Both Alex and Phoebe planned the methodology and did research on related work. Alex took the lead on writing the final paper and poster and retrieving accurate, well-controlled data from the dataset to help Phoebe train.

References

[1] anishathalye, Neural-Style Implementation. https://github.com/anishathalye/neural-style

[2] TensorFlow 1.13.1

[3] L. A. Gatys, A. S. Ecker, and M. Bethge. A neural algorithm of artistic style. *arXiv preprint arXiv:1508.06576*, 2015.

[4] L. A. Gatys, A. S. Ecker, and M. Bethge. Image style transfer using convolutional neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2414–2423, 2016.

[5] L. A. Gatys, M. Bethge, A. Hertzmann, and E. Shechtman. Preserving color in neural artistic style transfer. *arXiv preprint arXiv:1606.05897*, 2016.

[6] F. Luan, S. Paris, E. Shechtman, and K. Bala. Deep photo style transfer. In *CVPR*, 2017.

[7] Daan Wynen, Cordelia Schmid, and Julien Mairal. Unsupervised learning of artistic styles with archetypal style analysis. *arXiv preprint arXiv:1805.11155*, 2018.

[8] V. Dumoulin, J. Shlens, and M. Kudlur. A learned representation for artistic style. In *ICLR*, 2017.

[9] X. Huang and S. Belongie. Arbitrary style transfer in real-time with adaptive instance normalization. *arXiv preprint arXiv:1703.06868*, 2017.